

Exhibit 2

1

1 UNITED STATES DISTRICT COURT

2 SOUTHERN DISTRICT OF NEW YORK

3

4 THE NEW YORK TIMES COMPANY,

5 Plaintiff,

Civil Action No.:
1:23-cv-11195

6 vs.

7 MICROSOFT CORPORATION; OPENAI,
INC; OPENAI LP; OPENAI GP, LLC;
8 OPENAI, LLC; OPENAI OPCO LLC;
OPENAI GLOBAL LLC; OAI CORPORATION,
9 LLC; and OPENAI HOLDINGS, LLC;

10 Defendants.

-----/

11 (Caption continued on following page)

12

13

14 ***CONFIDENTIAL - ATTORNEY'S EYES ONLY***

15 VIDEO-RECORDED DEPOSITION OF MICHAEL TRINH

16 Wednesday, January 29, 2025

17

18

19

Stenographically reported by:

LORRIE L. MARCHANT, RMR, CRR, CCRR, CRC

California CSR No. 10523

Washington CSR No. 3318

Oregon CSR No. 19-0458

Texas CSR No. 11318

23

24 Job No.: 2025-971601

25

California Firm Registration No.: 179

58

1 Books2 was provided by legal counsel at OpenAI?

2 MR. GOLDBERG: Objection. Beyond the
3 scope. And that, as phrased, calls for privileged
4 communication.

5 If you know the answer that is not any
6 privileged communication, you can provide it.

7 THE WITNESS: The decision involved company
8 attorneys.

9 BY MR. STOLER:

10 Q. Do you know which attorneys?

11 A. Jason Kwon.

12 Q. How do you spell the last name?

13 A. K-W-O-N.

14 Q. Any others?

15 A. Che Chang.

16 Q. Any others aside from Jason and Che?

17 A. No, not that I can recall.

18 Q. And in the course of your investigation,
19 did you uncover any other datasets that were
20 deleted?

21 A. So I want to be precise here.

22 My understanding is -- so I -- in terms of
23 nomenclature, Books1 and Books2, I answered them in
24 the course of GPT-3.

25 I understand there was a similar dataset

1 involved for GPT-3.5. That was also part of that
2 decision. It's a subset of Books1 and Books2.

3 Q. What do you mean by "in the course of
4 GPT-3"?

5 A. You asked me for a dataset specific to the
6 training of GPT-3. That's what I meant.

7 Q. Okay. My question is, as to the training
8 of any of the OpenAI models, are you aware of any
9 training datasets that were deleted, aside from
10 Books1 and Books2?

11 A. What I'm trying to say is that Books- --
12 there was a subset of Books1 and Books2 used for the
13 training of ChatG- -- sorry -- GPT-3.5. That's
14 the -- that's the only other example I can think of.

15 Q. And do you have a name for that dataset?

16 A. I know it's a little confusing; but, no, I
17 don't.

18 Q. Is it all right if we call it the GPT-3.5
19 Books1 and Books2?

20 A. That works for me. I just wanted to make
21 sure that we were speaking precisely.

22 Q. Thank you.

23 All right. So it's your testimony that,
24 outside of Books1 and Books2, as well as the GPT-3.5
25 Books1 and Books2 datasets, OpenAI did not delete

1 large language model, are you referring to
2 exclusively to pretraining?

3 A. It could be called -- "pretraining" is
4 another term for referring to training, yes.

5 Q. All right. So my question is, are there
6 any datasets that were used for other phases of
7 training, such as fine-tuning, that OpenAI has
8 deleted?

9 A. I am not aware of any deletions of
10 fine-tuning datasets.

11 Q. How about the deletion of datasets used in
12 any other phase of developing a model at OpenAI?

13 A. I am not aware of any deletion. The
14 closest I can think of is a dataset that was
15 provided to us by a third party that mistakenly
16 contained personal information.

17 That dataset was segregated but not
18 deleted.

19 Q. And do you know why Books1 and Books2 were
20 deleted?

21 MR. GOLDBERG: Objection. Based on the
22 prior answers, that calls for privileged
23 information.

24 I instruct you not to answer that.

25 THE WITNESS: The decision was made with a

62

1 consultation with the company attorneys, and I can't
2 answer further without revealing privileged
3 information.

4 BY MR. STOLER:

5 Q. All right. So you mentioned just now
6 processing datasets.

7 Do you recall that?

8 A. I do.

9 Q. What do you mean by "processing datasets"?

10 A. There's a couple steps that are taken to
11 put data in a format that is -- that is functionally
12 capable of training a model.

13 Q. And does -- what does that process -- what
14 are examples of that type of processing?

15 A. One component is tokenization. And --
16 yeah.

17 Q. Anything else?

18 A. One process is filtering.

19 Q. Does OpenAI store preprocessed data?

20 A. I don't know.

21 Q. Who would know?

22 A. I would have to check with Ian.

23 Q. And for each model trained by OpenAI, does
24 OpenAI store a complete set of the training data
25 used to train that model?

211

1 STENOGRAPHER'S CERTIFICATE

2 I, LORRIE L. MARCHANT, Certified Shorthand
3 Reporter, Certificate No. 10523, for the State of
4 California, hereby certify that MICHAEL TRINH was
5 by me duly sworn/affirmed to testify to the truth,
6 the whole truth and nothing but the truth in the
7 within-entitled cause; that said deposition was
8 taken at the time and place herein named; that the
9 deposition is a true record of the witness's
10 testimony as reported to the best of my ability by
11 me, a duly certified shorthand reporter and a
12 disinterested person, and was thereafter transcribed
13 under my direction into typewriting by computer;
14 that request [] was [x] was not made to read and
15 correct said deposition.

16 I further certify that I am not interested
17 in the outcome of said action, nor connected with,
18 nor related to any of the parties in said action,
19 nor to their respective counsel.

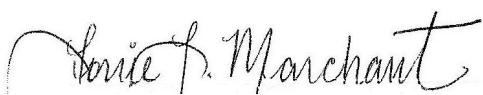
20 IN WITNESS WHEREOF, I have hereunto set my
21 hand this 3rd day of February, 2025.

22

23

24

25



LORRIE L. MARCHANT, RMR, CRR, CCRR, CRC
Stenographic Certified Shorthand Reporter #10523